

## CLAIMS

That which is claimed is:

1. A method for overlaying gene- or protein-related data on chromosome maps, said method comprising the steps of:

importing arbitrary gene- or protein-related data having identifiers for determining genetic loci of genes to which said arbitrary gene-related data are associated;

reading the identifiers;

matching the identifiers with predefined identifiers on at least one of the chromosome maps; and

displaying the arbitrary gene- or protein related data adjacent positions on the at least one chromosome map where the genes associated with the respective arbitrary gene- or protein-related data are located, wherein said importing, reading, matching and displaying are all automated steps.

2. The method of claim 1, further comprising interactive selection by a user of at least one data type to be displayed during said displaying.

3. The method of claim 1, further comprising spatially grouping said gene- or protein-related data to correspond to spatial groupings of said associated genes on said at least one chromosome map.

4. The method of claim 1, further comprising compressing said gene- or protein-related data when required to display said gene- or protein-related data in an area in which all of the gene- or protein-related data cannot be discretely displayed.

5. The method of claim 1, further comprising zooming at least one of said gene- or protein-related data and said at least one chromosome map to display an enlarged view of additional detail relevant to a zoomed area.

6. The method of claim 1, further comprising querying and cutting information on the display that a user is not interested in viewing.

7. The method of claim 1, wherein said at least one chromosome map comprises a plurality of chromosome maps, said method further comprising maintaining focus and context of at least a portion of the display of said chromosome maps and gene- or protein-related data.

8. The method of claim 7, further comprising displaying a high level view of all of said chromosome maps and gene- or protein-related data, a mid-level view displaying a magnified view of a selected portion of said high level view, and a detailed view displaying expanded, detailed information characterizing a selected portion of said mid-level view.

9. The method of claim 8, wherein said high-level view, mid-level view and detailed view are all interlinked so that changing one view automatically changes the other two views in the same way, substantially simultaneously.

10. The method of claim 1, further comprising displaying tooltips to display additional details relative to a selected portion of the display.

11. The method of claim 1, further comprising displaying popup dialogs to display additional details relative to a selected portion of the display.

12. The method of claim 1, further comprising accessing an external source of information relative to the data displayed, matching at least one of said identifiers with specific information in said external source; and displaying said specific information relative to said gene- or protein-related data associated with said at least one identifier.

13. The method of claim 1, wherein said identifiers of said arbitrary gene- or protein-related data are selected from published gene identifiers and symbols.

14. The method of claim 13, wherein said published gene identifiers and symbols are selected from at least one of GenBank accession numbers, RefSeq accession numbers, UniGene Cluster ID's, UniGene ID's, official standard gene names, LocusLink ID, SwissProt ID's, and Protein Information Resource (PIR) ID's.

15. The method of claim 1, wherein said matching comprises providing a relational database which stores a set of cross-referenced tables for matching said identifiers with said predefined identifiers, and as the identifiers are read, they are matched with said predefined identifiers in the cross-referenced tables through standard database queries.

16. The method of claim 1, wherein said arbitrary gene- or protein-related data comprises an expression matrix.

17. The method of claim 16, wherein said arbitrary gene- or protein-related data comprises a plurality of expression matrices.

18. The method of claim 1, wherein said arbitrary gene- or protein-related data comprises a matrix of at least one microarray of gene expression data, wherein each row of the matrix is associated with a particular gene, and wherein said matching comprises reordering and spatial grouping of the rows based on matching the identifiers to the predefined identifiers.

19. The method of claim 18, wherein a visualization of the matrix resultant from said displaying comprises a heat map.

20. The method of claim 1, further comprising statistically assessing co-location values and displaying assessed co-location statistical significance ~~values-assessed~~ along side said arbitrary gene-related data.

21. The method of claim 1, further comprising the steps of:  
selecting additional information characterizing said arbitrary gene- or protein-related data; and  
displaying said additional information along side of said display of the arbitrary gene- or protein-related data and positioned relative to the respective locations on the chromosome map of the respective genes characterized by said arbitrary gene- or protein-related data.
22. The method of claim 21, wherein said additional information comprises annotations.
23. The method of claim 22, wherein said annotations comprise gene ontology annotations.
24. The method of claim 21, wherein said additional information is selected from the group consisting of CGH data, protein levels, relevance scores and relevance densities.
25. The method of claim 22, wherein said arbitrary gene- or protein-related data is displayed in matrix format and said additional information is displayed in at least one additional matrix.
26. The method of claim 21, wherein said arbitrary gene- or protein-related data is displayed in scatter plot format.
27. The method of claim 1, wherein said arbitrary gene- or protein- related data is imported from a plurality of experiments.
28. The method of claim 27, wherein said arbitrary gene- or protein- related data is displayed with regard to each of the plurality of experiments on a single display.
29. The method of claim 21, wherein said additional information includes at

least one of annotations, cellular localization of the genetic material, cluster data, and statistical data.

30. The method of claim 18, further comprising calculating row vectors of the values in the rows of the matrix; using an auxiliary process to obtain cluster data for said row vectors; and displaying said cluster data along side said display of said arbitrary gene- or protein-related data.

31. The method of claim 30, wherein said matrix comprises a heat map, and wherein said cluster data and said arbitrary gene- or protein-related data are displayed with color coding.

32. The method of claim 30, wherein said cluster data is displayed in a single column adjacent each matrix of gene- or protein-related data.

33. The method of claim 30, wherein said cluster data is displayed in a multi-column matrix adjacent each matrix of gene- or protein-related data, respectively.

34. The method of claim 1, wherein said arbitrary gene- or protein-related data comprises a matrix of at least one microarray of gene expression data, wherein each row of the matrix is associated with a particular gene, and wherein each column of the matrix is associated with a microarray experiment, wherein a portion of the total number of columns are associated with experiments taken from normal, healthy tissue, and another portion of the total number of columns are associated with experiments taken from tissue exhibiting a known abnormality, said method further comprising dividing the matrix into two smaller matrices with a first matrix containing the columns associated with normal experiments and a second matrix containing the columns associated with abnormal experiments, and wherein said matching and displaying are performed with regard to both first and second matrices.

35. The method of claim 34, wherein the first and second matrices are displayed in color coding as heat maps.

36. The method of claim 34, further comprising calculating a relevance score for at least one row of the matrices by comparing expression values in the first matrix with expression values in the second matrix, and displaying at least one calculated relevance score along side the row to which each pertains.

37. The method of claim 36 wherein said calculating is interactively initiated via a user interface.

38. The method of claim 36, wherein the relevance score comprises a "p value" and the relevance score is displayed as a valued calculated by  $(-\log p \text{ value})$ .

39. The method of claim 36, wherein a plurality of relevance scores are calculated and displayed as a line map.

40. The method of claim 36, wherein a plurality of relevance scores are calculated and displayed in color-coding as a heat map.

41. The method of claim 36, wherein relevance scores are calculated and displayed in a binary code.

42. The method of claim 36, wherein a plurality of relevance scores are calculated, said method further comprising defining a relevance density score based upon distances between genetic locations and relevance scores, and identifying chromosomal locations containing relevance density scores greater than or equal to the defined relevance density score.

43. The method of claim 36, wherein a plurality of relevance scores are calculated, said method further comprising filtering the relevance scores by setting at least one relevance score limit value and displaying only those relevance scores which are greater than or equal to at least one relevance score limit value.

44. The method of claim 34, further comprising matching chromosomal copy number abnormality data with the gene-related data identifiers, and displaying the

chromosomal copy number abnormality data along side the gene-related data to which each is matched.

45. The method of claim 44, wherein the chromosomal copy number abnormality data is displayed in third and fourth matrices, wherein each value in the third matrix is matched with the expression value in the first matrix having the same row and column location, and wherein each value in the fourth matrix is matched with the expression value in the second matrix having the same row and column location.

46. The method of claim 44, wherein the chromosomal copy number abnormality data is provided in columns interlaced with the columns of expression data in the first and second matrices.

47. The method of claim 44, wherein the chromosomal copy number abnormality is displayed in color-coding, as one or more heat maps.

48. The method of claim 44, wherein the chromosomal copy number abnormality data is displayed as one or more line maps.

49. The method of claim 45, further comprising calculating a relevance score for at least one row of the chromosomal copy number abnormality data by comparing chromosomal copy number abnormality values in the third matrix with chromosomal copy number abnormality values in the fourth matrix, and displaying at least one calculated relevance score along side the row to which each pertains.

50. The method of claim 49, wherein the relevance score comprises a "p value" and the relevance score is displayed as a valued calculated by  $(-\log p \text{ value})$ .

51. The method of claim 49, wherein a plurality of relevance scores are calculated and displayed as a line map.

52. The method of claim 49, wherein a plurality of relevance scores are calculated and displayed in color-coding as a heat map.

53. The method of claim 49, wherein a plurality of relevance scores are calculated, said method further comprising defining a relevance density score based upon distances between genetic locations and relevance scores, and identifying chromosomal locations containing relevance density scores greater than or equal to the defined relevance density score.

54. The method of claim 49, wherein a plurality of relevance scores are calculated, said method further comprising filtering the relevance scores by setting at least one relevance score limit value and displaying only those relevance scores which are meet or exceed at least one relevance score limit value.

55. The method of claim 1, further comprising the steps of:

selecting additional information related to one or more genes characterized by said arbitrary gene- or protein-related data; and

displaying said additional information along side of said display of the arbitrary gene- or protein-related data and positioned relative to the respective locations on the chromosome map of the respective genes characterized by said arbitrary gene- or protein-related data.

56. The method of claim 55, wherein said additional information comprise at least one of polymorphism measurements, annotations, transcription factor binding sites, RNA expression values, allele information, alternative exon splicing data, mapping of CGH gene amplification/deletions, and protein abundance.

57. A method for overlaying gene-related data on chromosome maps, said method comprising the steps of:

importing gene expression data in the form of at least one heat map having identifiers for determining genetic loci of genes to which rows of said expression data are associated;

reading the identifiers;

matching the identifiers with predefined identifiers on at least one of the chromosome maps;



reordering the rows of each said heat map, based on said matching, so that said rows occur in the order of location of the associated genes on the one or more chromosome maps; and

displaying the reordered rows adjacent the at least one chromosome map where the genes associated with the expression data are located, wherein said importing, reading, matching and displaying are all automated steps.

58. The method of claim 57, further comprising spatial grouping said rows to correspond to spatial groupings on the one or more chromosome maps of the respective genes associated with the expression data.

59. The method of claim 57, wherein the reordered rows are displayed with even spacing, said method further comprising overlaying mapping lines that connect said rows of expression data with spatial locations on the at least one chromosome map of the respective genes associated with the rows.

60. The method of claim 57, further comprising statistically assessing co-location values of said data and displaying assessed co-location statistical significance along side said arbitrary gene-related data.

61. The method of claim 57, further comprising the steps of:

selecting additional information relating to one or more genes associated with said expression data; and

displaying said additional information along side of said display of the expression data and positioned relative to the respective locations on the at least one chromosome map of the respective genes characterized by said expression data.

62. The method of claim 57, further comprising calculating row vectors of the expression values in the rows of the heat map; using an auxiliary process to obtain cluster data for said row vectors; and displaying said cluster data along side said display of said expression data and positioned relative to the respective locations on the at least one chromosome map of the respective genes characterized by said expression data.

63. The method of claim 62, wherein said cluster data and said expression values are displayed with color coding.

64. The method of claim 62, wherein said cluster data is displayed in a single column relative to each matrix of expression data.

65. The method of claim 57, further comprising calculating a relevance score for at least one row of the heat map and displaying at least one calculated relevance score in a position relative to the row to which each pertains.

66. The method of claim 65, wherein a plurality of relevance scores are calculated, said method further comprising defining a relevance density score based upon distances between genetic locations and relevance scores, and identifying chromosomal locations containing relevance density scores greater than or equal to the defined relevance density score.

67. The method of claim 65, wherein a plurality of relevance scores are calculated, said method further comprising filtering the relevance scores by setting at least one relevance score limit value and displaying only those relevance scores which meet or exceed the at least one relevance score limit value.

68. The method of claim 57, further comprising matching chromosomal copy number abnormality data with the expression data identifiers, and displaying the chromosomal copy number abnormality data relative to the display of the expression data.

69. A system for displaying visualizations of gene-related data on chromosomal graphic schemes, said system comprising:

- means for automatically generating chromosome maps;
- means for automatically inputting gene- or protein-related data;
- means for automatically reading identifiers associating gene- or protein-related data with genes which said gene- or protein-related data are associated with;

means for automatically matching said identifiers with locations on at least one chromosome map on which said genes are located;

means for automatically ordering said gene- or protein-related data to correspond to respective locations of said associated genes on said at least one chromosome map; and

means for automatically displaying said gene- or protein-related data relative to the locations of the genes associated with said gene- or protein-related data, respectively.

70. The system of claim 69, further comprising means for spatially grouping said reordered gene- or protein-related data to correspond to spatial groupings of said associated genes on said at least one chromosome map.

71. The system of claim 69, further comprising means for compressing said gene- or protein-related data when required to display said gene- or protein-related data in an area in which all of the gene- or protein-related data cannot be discretely displayed.

72. The system of claim 69, further comprising means for zooming at least one of said gene- or protein-related data and said at least one chromosome map to display an enlarged view of additional detail relevant to a zoomed area.

73. The system of claim 69, further comprising means for querying and cutting information on the display that a user is not interested in viewing.

74. The system of claim 69, wherein said at least one chromosome map comprises a plurality of chromosome maps, said system further comprising means for maintaining focus and context of at least a portion of the display of said chromosome maps and gene- or protein-related data.

75. The system of claim 73, wherein said system comprises means for displaying a high level view of all of said chromosome maps and gene- or protein-related data, a mid-level view displaying a magnified view of a selected portion of said

high level view, and a detailed view displaying expanded, detailed information characterizing a selected portion of said mid-level view.

76. The system of claim 75, wherein said high-level view, mid-level view and detailed view are all interlinked so that changing one view automatically changes the other two views in the same way, substantially simultaneously.

77. The system of claim 69, further comprising means for displaying tooltips to display additional details relative to a selected portion of the display.

78. The system of claim 69, further comprising means for displaying popup dialogs to display additional details relative to a selected portion of the display.

79. The system of claim 69, further comprising means for accessing an external source of information relative to the data displayed, means for matching at least one of said identifiers with specific information in said external source; and means for displaying said specific information relative to said gene- or protein-related data associated with said at least one identifier.